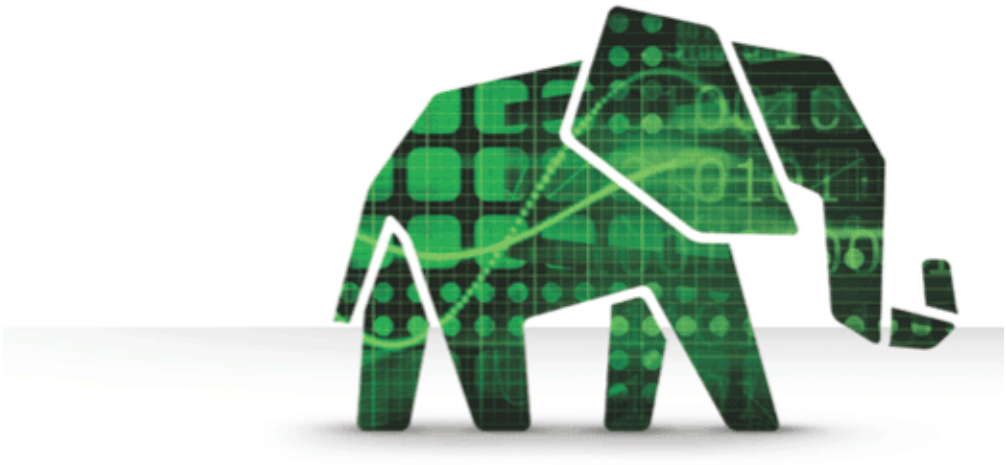
Hortonworks Architecting the Future of Big Data

Hortonworks Technical Preview for Apache Slider

Released: May 30, 2014

Revision: 0.30.002



Architecting the Future of Big Data

Welcome to **Hortonworks Apache Slider Technical Preview**. This Technical Preview provides early access to upcoming features, letting you test and review during the development process. These features are considered under development.

These features are not intended for use in your production systems and are not supported by Hortonworks but your feedback is greatly appreciated.

Have fun and please send feedback on the Hortonworks Community forums
http://hortonworks.com/community/forums/forum/slider/

Hortonworks

Introduction

Apache Slider is a project in incubation at the Apache Software Foundation with the goal of making it possible and easy to deploy existing applications onto a YARN cluster. The following provides the steps required for setting up a cluster and deploying a YARN hosted application using the **Apache Slider Technical Preview**.

System Requirements

The Slider deployment has the following minimum system requirements:

- Hortonworks Data Platform 2.1
- Required Services: HDFS, YARN and ZooKeeper
- Oracle JDK 1.7 (64-bit)

Operating Systems

- 64-bit Red Hat Enterprise Linux (RHEL) 6
- 64-bit CentOS 6
- 64-bit Oracle Linux 6

Setup the Cluster

After installing your cluster (using Apache Ambari or other means – refer to http://docs.hortonworks.com/) with the services listed above, modify your YARN configuration to allow for multiple containers on a single host. In `yarn-site.xml` make the following modifications:

Property	Value
`yarn.scheduler.minimum-allocation-mb`	`1`
`yarn.nodemanager.delete.debug-delay-sec`	`>= 3600 (to retain for an hour)`
`yarn.nodemanager.pmem-check-enabled`	`false`
`yarn.nodemanager.vmem-check-enabled`	`false`

There are other options detailed in the Troubleshooting file available at:

https://github.com/apache/incubator-slider/blob/master/src/site/markdown/troubleshooting.md

Download Slider 0.30 Package

The Slider package is available at:

http://public-repo-1.hortonworks.com/slider/0.30/slider-0.30-all.tar

Install Slider

Complete the following steps to expand and install Slider:

- `mkdir <slider-install-dir>`
- `cd <slider-install-dir>`
- Log in as the 'yarn' user (assuming this is a host associated with the installed cluster). *This assumes that all apps are being run as 'yarn' user. Any other user can be used to run the apps - ensure that file permission is granted as required.*
- Expand the tar file: `tar -xvf slider-0.30-all.tar`
- Browse to the Slider directory: `cd slider-0.30/bin`
- `export PATH=$PATH:/usr/jdk64/jdk1.7.0_45/bin` (or the path to the JDK bin directory)
- Modify Slider configuration file `<slider-install-dir>/slider-0.30/conf/slider-client.xml` to add the following properties:

```
<property>
  <name>yarn.application.classpath</name>
<value>/etc/hadoop/conf,/usr/lib/hadoop/*,/usr/lib/hadoop/lib/*,/usr/lib/had
oop-hdfs/*,/usr/lib/hadoop-hdfs/lib/*,/usr/lib/hadoop-
yarn/*,/usr/lib/hadoop-yarn/lib/*,/usr/lib/hadoop-
mapreduce/*,/usr/lib/hadoop-mapreduce/lib/*</value>
</property>
<property>
  <name>slider.zookeeper.quorum</name>
  <value>yourZooKeeperHost:port</value>
</property>
```

In addition, specify the scheduler and HDFS addresses as follows:

```
<property>
  <name>yarn.resourcemanager.address</name>
  <value>yourResourceManagerHost:8050</value>
</property>
<property>
  <name>yarn.resourcemanager.scheduler.address</name>
  <value>yourResourceManagerHost:8030</value>
</property>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://yourNameNodeHost:8020</value>
</property>
```

- Execute: `<slider-install-dir>/slider-0.30/bin/slider version`
- Ensure there are no errors and you can see "`Compiled against Hadoop 2.4.0`"

Deploy Slider Resources

Ensure that all file folders are accessible to the user creating the application instance. The example assumes “yarn” to be that user.

Create HDFS root folder for Slider

Complete the following steps to create the Slider root folder with the appropriate permissions:

```
su hdfs
hdfs dfs -mkdir /slider
hdfs dfs -chown yarn:hdfs /slider
hdfs dfs -mkdir /user/yarn
hdfs dfs -chown yarn:hdfs /user/yarn
```

Load Slider Agent

```
su yarn
hdfs dfs -mkdir /slider/agent
hdfs dfs -mkdir /slider/agent/conf
hdfs dfs -copyFromLocal <slider-install-dir>/slider-0.30/agent/slider-
agent.tar.gz /slider/agent
```

Create and deploy Slider Agent configuration

Create an agent config file (`agent.ini`) based on the sample available at:

```
<slider-install-dir>/slider-0.30/agent/conf/agent.ini
```

The sample `agent.ini` file can be used as is (see below). Some of the parameters of interest are:

- log_level = INFO or DEBUG, to control the verbosity of log
- app_log_dir = the relative location of the application log file
- log_dir = the relative location of the agent and command log file

```
[server]
hostname=localhost
port=8440
secured_port=8441
check_path=/ws/v1/slider/agents/
register_path=/ws/v1/slider/agents/{name}/register
heartbeat_path=/ws/v1/slider/agents/{name}/heartbeat

[agent]
app_pkg_dir=app/definition
app_install_dir=app/install
app_run_dir=app/run
app_task_dir=app/command-log
app_log_dir=app/log
```

```
app_tmp_dir=app/tmp
log_dir=infra/log
run_dir=infra/run
version_file=infra/version
log_level=INFO

[python]

[command]
max_retries=2
sleep_between_retries=1

[security]

[heartbeat]
state_interval=6
log_lines_count=300
```

Once created, deploy the agent.ini file to HDFS:

```
su yarn
hdfs dfs -copyFromLocal agent.ini /slider/agent/conf
```

Download Sample Application Packages

There are three sample application packages available for download to use with Slider:

Application	Version	URL
Apache HBase	0.96.0	http://public-repo-1.hortonworks.com/slider/0.30/apps/hbase_v096.zip
Apache Storm	0.9.1	http://public-repo-1.hortonworks.com/slider/0.30/apps/storm_v091.zip
Apache Accumulo	1.5.1	http://public-repo-1.hortonworks.com/slider/0.30/apps/accumulo_v151.zip

Download the packages and deploy one of these sample applications to YARN via Slider using the following steps.

Install, Configure, Start and Verify Sample Application

- Load Sample Application Package
- Create Application Specifications
- Start the Application
- Manage the Application Lifecycle
- Application Registry

Load Sample Application Package

```
hdfs dfs -copyFromLocal <sample-application-package> /slider
```

If necessary, create HDFS folders needed by the application. For example, HBase requires the following HDFS-based setup:

```
su hdfs
hdfs dfs -mkdir /apps
hdfs dfs -mkdir /apps/hbase
hdfs dfs -chown yarn:hdfs /apps/hbase
```

Create Application Specifications

Configuring a Slider application requires that you create two files: the *Resource Specification*, and the *Application Configuration*. Guidelines for creating these files follow.

Note: There are sample Resource Specifications (`resources.json`) and Application Configuration (appConfig.json) files in the Appendix and also in the root directory of the Sample Applications packages (e.g. `/hbase_v096.zip/resources.json` and `/hbase_v096.zip/appConfig.json`).

Resource Specification

The **Resource Specification**, or **Resource Spec**, file provides Slider the following deployment requirements for the application package:

- What components to deploy
- How many components to deploy
- Memory and CPU requirements for YARN

For example, in HBase, the components are ***master*** and ***worker*** -- the former hosting the **HBase Master** and the latter hosting **HBase RegionServers.** Put these facts in the Resource Spec file, named `resources.json`.

Sample Resource Spec files are available in the Appendix:

- Appendix A: Storm Sample Resource Specification
- Appendix B: HBase Sample Resource Specification

Store the Resource Spec file on your local disk (e.g. `/tmp/resources.json`).

Application Configuration

The **Application Configuration**, or **App Config**, file provides Slider with all required application- and component-specific configuration parameters, such as heap sizes of the JVMs. The App Config also defines configuration details specific to the application and component instances.

Sample App Configs are available in the Appendix:

- Appendix A: Storm Sample Application Configuration

- Appendix B: HBase Sample Application Configuration

Start the Application

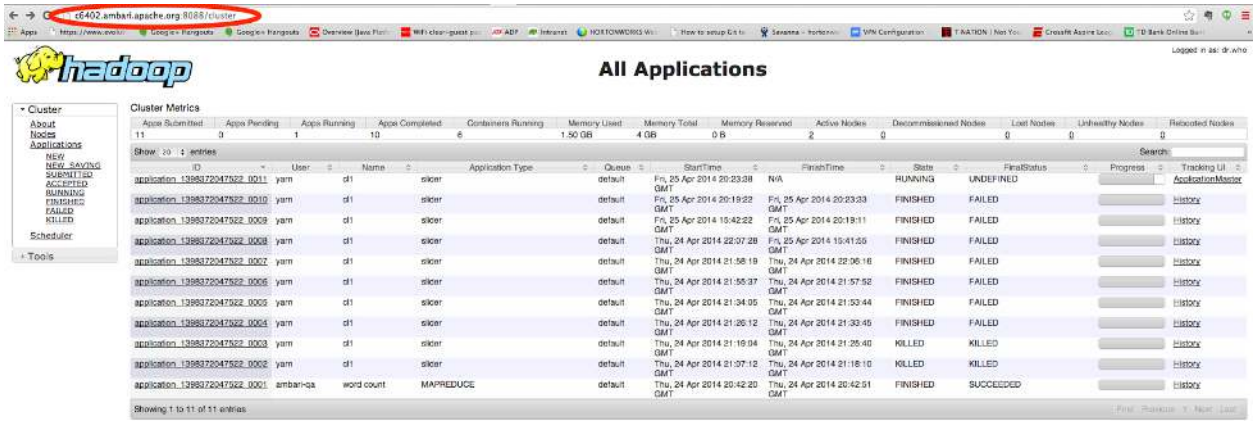
After installing Slider and creating the specification files, the application by leveraging the Slider Command Line Interface (CLI).

- Change directory to the "bin" directory under the slider installation
  `cd <slider-install-dir>/slider-0.30/bin`
- Execute the following command:

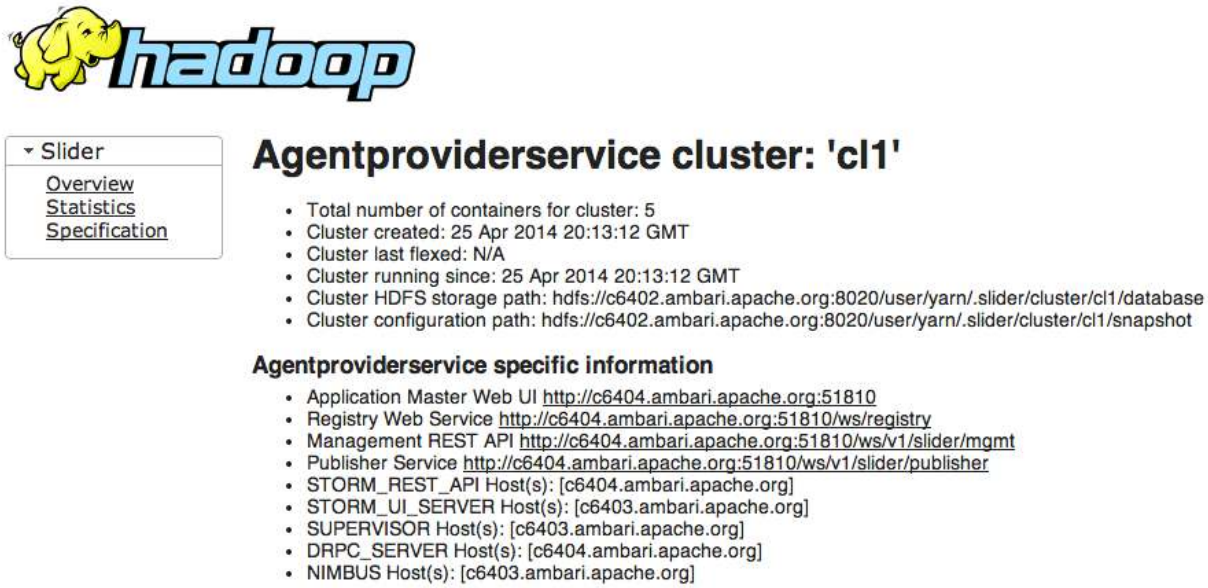
```
./slider create cl1 --image hdfs://yourNameNodeHost:8020/slider/agent/slider-agent.tar.gz --template appConfig.json --resources resources.json
```

Verify the Application

Verify that Slider launches successfully using the YARN Resource Manager Web UI. In most instances, this UI is accessible via a web browser at port 8088 of the Resource Manager Host:



The specific information for the running application is accessible via the "ApplicationMaster" link that can be seen in the far right column of the row associated with the running application (probably the top row):



Manage the Application Lifecycle

Once started, applications can be frozen/stopped, thawed/restarted, and destroyed/removed as follows:

Freeze

```
./slider freeze cl1 --manager yourResourceManagerHost:8050 --filesystem
hdfs://yourNameNodeHost:8020
```

Thaw

```
./slider thaw cl1 --manager yourResourceManagerHost:8050 --filesystem
hdfs://yourNameNodeHost:8020
```

Destroy

```
./slider destroy cl1 --manager yourResourceManagerHost:8050 --filesystem
hdfs://yourNameNodeHost:8020
```

Application Registry

Each application publishes several artifacts that can be used by an application administrator or application client. Typical data published includes the applied configuration, links to application JMX endpoint or monitoring UI and log folders.

All published data is available at the publisher endpoint that is hosted by the Slider Application Master. An example publisher endpoint is:

```
http://c6401.ambari.apache.org:47457/ws/v1/slider/publisher
```

From this endpoint, you can access configuration information published by the application.

Publisher URI	Description
`{publisher-endpoint}/slider/quicklinks`	Named URLs that app publishes
`{publisher-endpoint}/slider/logfolders`	Log folders for the app components (YARN should be configured to retain logs)
`{publisher-endpoint}/slider/storm-site`	Applied configs by the app (e.g. storm-site, hbase-site)

Example output from `/slider/quicklinks`:


```
{
  "description": "QuickLinks",
  "entries": {
    "org.apache.slider.jmx": "http://c6401.ambari.apache.org:50154/api/cluster/summary",
    "org.apache.slider.metrics": "http://c6401.ambari.apache.org/cgi-bin/rrd.py?c=Application2",
    "org.apache.slider.monitor": "http://c6401.ambari.apache.org:41806",
    "org.apache.slider.ganglia": "http://c6401.ambari.apache.org/ganglia?c=Application2"
    },
  "updated": 0,
  "empty": false
}
```

Known Issues and Limitations

At the time of this release, there are no known issues for the **Slider Technical Preview** but you can visit the forum for the latest discussions on issues:

http://hortonworks.com/community/forums/forum/slider

Troubleshooting

Please refer to the Troubleshooting file available at:

https://github.com/apache/incubator-slider/blob/master/src/site/markdown/troubleshooting.md

Further Reading

Additional **Slider** information is available here:

- http://slider.incubator.apache.org/
- https://github.com/apache/incubator-slider/blob/master/src/site/markdown/slider_specs/index.md
- http://incubator.apache.org/projects/slider.html
- http://hortonworks.com/community/forums/forum/slider

Hortonworks

Appendix A: Apache Storm Sample Application Specs

Storm Resource Specification Sample

```
{
    "schema" : "http://example.org/specification/v2.0.0",
    "metadata" : {
    },
    "global" : {
    },
    "components" : {
        "slider-appmaster" : {
        },
        "NIMBUS" : {
            "role.priority" : "1",
            "component.instances" : "1"
        },
        "STORM_REST_API" : {
            "role.priority" : "2",
            "component.instances" : "1"
        },
        "STORM_UI_SERVER" : {
            "role.priority" : "3",
            "component.instances" : "1"
        },
        "DRPC_SERVER" : {
            "role.priority" : "4",
            "component.instances" : "1"
        },
        "SUPERVISOR" : {
            "role.priority" : "5",
            "component.instances" : "1"
        }
    }
}
```

Storm Application Configuration Sample

```
{
    "schema" : "http://example.org/specification/v2.0.0",
    "metadata" : {
    },
    "global" : {
        "A site property for type XYZ with name AA": "its value",
        "site.XYZ.AA": "Value",
        "site.hbase-site.hbase.regionserver.port": "0",
        "site.core-site.fs.defaultFS": "${NN_URI}",
        "Using a well known keyword": "Such as NN_HOST for name node host",
        "site.hdfs-site.dfs.namenode.http-address": "${NN_HOST}:50070",
        "a global property used by app scripts": "not affiliated with any site-xml",
        "site.global.app_user": "yarn",
        "Another example of available keywords": "Such as AGENT_LOG_ROOT",
        "site.global.app_log_dir": "${AGENT_LOG_ROOT}/app/log",
        "site.global.app_pid_dir": "${AGENT_WORK_ROOT}/app/run",
```

```
    }
}
```

Appendix B: Apache HBase Sample Application Specs

HBase Resource Specification Sample

```
{
  "schema": "http://example.org/specification/v2.0.0",
  "metadata": {
  },
  "global": {
  },
  "components": {
    "HBASE_MASTER": {
      "role.priority": "1",
      "component.instances": "1"
    },
    "slider-appmaster": {
    },
    "HBASE_REGIONSERVER": {
      "role.priority": "2",
      "component.instances": "1"
    }
  }
}
```

HBase Application Configuration Sample

```
{
  "schema": "http://example.org/specification/v2.0.0",
  "metadata": {
  },
  "global": {
    "agent.conf": "/slider/agent/conf/agent.ini",
    "application.def": "/slider/hbase_v096.zip",
    "config_types": "core-site,hdfs-site,hbase-site",
    "java_home": "/usr/jdk64/jdk1.7.0_45",
    "package_list": "files/hbase-0.96.1-hadoop2-bin.tar.gz",
    "site.global.app_user": "yarn",
    "site.global.app_log_dir": "${AGENT_LOG_ROOT}/app/log",
    "site.global.app_pid_dir": "${AGENT_WORK_ROOT}/app/run",
    "site.global.app_root": "${AGENT_WORK_ROOT}/app/install/hbase-0.96.1-hadoop2",
    "site.global.app_install_dir": "${AGENT_WORK_ROOT}/app/install",
    "site.global.hbase_master_heapsize": "1024m",
    "site.global.hbase_regionserver_heapsize": "1024m",
    "site.global.user_group": "hadoop",
    "site.global.security_enabled": "false",
    "site.global.ganglia_server_host": "${NN_HOST}",
    "site.global.ganglia_server_port": "8667",
    "site.global.ganglia_server_id": "Application1",
    "site.hbase-site.hbase.hstore.flush.retries.number": "120",
    "site.hbase-site.hbase.client.keyvalue.maxsize": "10485760",
```

```
    "site.hbase-site.hbase.hstore.compactionThreshold": "3",
    "site.hbase-site.hbase.rootdir": "${NN_URI}/apps/hbase/data",
    "site.hbase-site.hbase.stagingdir": "${NN_URI}/apps/hbase/staging",
    "site.hbase-site.hbase.regionserver.handler.count": "60",
    "site.hbase-site.hbase.regionserver.global.memstore.lowerLimit": "0.38",
    "site.hbase-site.hbase.hregion.memstore.block.multiplier": "2",
    "site.hbase-site.hbase.hregion.memstore.flush.size": "134217728",
    "site.hbase-site.hbase.superuser": "yarn",
    "site.hbase-site.hbase.zookeeper.property.clientPort": "2181",
    "site.hbase-site.hbase.regionserver.global.memstore.upperLimit": "0.4",
    "site.hbase-site.zookeeper.session.timeout": "30000",
    "site.hbase-site.hbase.tmp.dir": "${AGENT_WORK_ROOT}/work/app/tmp",
    "site.hbase-site.hbase.local.dir": "${hbase.tmp.dir}/local",
    "site.hbase-site.hbase.hregion.max.filesize": "10737418240",
    "site.hbase-site.hfile.block.cache.size": "0.40",
    "site.hbase-site.hbase.security.authentication": "simple",
    "site.hbase-site.hbase.defaults.for.version.skip": "true",
    "site.hbase-site.hbase.zookeeper.quorum": "${ZK_HOST}",
    "site.hbase-site.zookeeper.znode.parent": "/hbase-unsecure",
    "site.hbase-site.hbase.hstore.blockingStoreFiles": "10",
    "site.hbase-site.hbase.hregion.majorcompaction": "86400000",
    "site.hbase-site.hbase.security.authorization": "false",
    "site.hbase-site.hbase.cluster.distributed": "true",
    "site.hbase-site.hbase.hregion.memstore.mslab.enabled": "true",
    "site.hbase-site.hbase.client.scanner.caching": "100",
    "site.hbase-site.hbase.zookeeper.useMulti": "true",
    "site.hbase-site.hbase.regionserver.info.port": "0",
    "site.hbase-site.hbase.master.info.port": "${HBASE_MASTER.ALLOCATED_PORT}",
    "site.hbase-site.hbase.regionserver.port": "0",
    "site.core-site.fs.defaultFS": "${NN_URI}",
    "site.hdfs-site.dfs.namenode.https-address": "${NN_HOST}:50470",
    "site.hdfs-site.dfs.namenode.http-address": "${NN_HOST}:50070"
  },
  "components": {
    "HBASE_MASTER": {
    },
    "slider-appmaster": {
      "jvm.heapsize": "256M"
    },
    "HBASE_REGIONSERVER": {
    }
  }
}
```

About Hortonworks

Hortonworks is a leading commercial vendor of Apache Hadoop, the preeminent open source platform for storing, managing and analyzing big data. Hortonworks Data Platform provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy big data solutions. Hortonworks is the trusted source for information on Hadoop, and together with the Apache community, Hortonworks is making Hadoop easier to install, manage and use. Hortonworks provides technical support, training & certification programs for enterprises, systems integrators & technology vendors.

3460 W. Bayshore Rd.
Palo Alto, CA 94303 USA

US: 1.855.846.7866
International: 1.408.916.4121
www.hortonworks.com